

Putting a National Portal for Undergraduate Theses into Production

Peter Linde¹, Carin Björklund², Aina Svensson³

¹ Blekinge Institute of Technology; SE-371 79 Karlskrona, Sweden, peter.linde@bth.se

² Göteborg University Library, Box 222, SE-405 30 Göteborg, Carin.Bjorklund@ub.gu.se

³ Swedish University of Agricultural Sciences, Box 7071, SE-750 07 Uppsala, aina.svensson@bibul.slu.se

Abstract

This paper discusses processes and experiences gained from creating a national portal (Uppsök) for Swedish undergraduate theses, using a common metadata model and set structure with agreements on semantics on top of OAI-PMH and harvesting from several data providers into a central service provider at the Swedish Royal Library.

1 Introduction

In Sweden there are about 40 universities and university colleges engaged in different stages of developing strategies and solutions for their own institutions' publications. The tasks are being carried out by different levels of staff, from the decision makers to enthusiasts, usually in collaboration with the library. As an increasing number of documents are published and distributed electronically, the need for coordination and promotion of standards in this area is vital.

1.1 Coordination Of Electronic Publishing At Swedish Universities And University Colleges

As a means of support and promotion of standardisation surrounding open institutional full text archives, several activities in Sweden have been carried out within a national project, called SVEP (SVEP project homepage, 2005). Based on a merger of project proposals submitted to BIBSAM in 2003 (the Swedish Royal Library's Department for National Co-ordination and Development) (BIBSAM, 2005), SVEP was created with core partners from nine Swedish universities and the Royal Library.

The SVEP project's objectives are to coordinate and scale up the development of e-publishing within the Swedish higher education sector, to promote efficiency and sustainability, and to avoid duplication of efforts. SVEP is realised through five work packages: 1. Interoperability, harmonisation of descriptions and description formats for electronically published scientific publications; 2. Preservation - developing workflow and technical solutions for archiving electronic publications; 3. Advice and dissemination of e-publishing, with an overview of institutional repository software which supports OAI; 4. Workshops - a series of workshops within the SVEP project; and 5. An undergraduate theses service - building a national portal for Swedish undergraduate theses and diploma work using OAI-PMH (OAI, 2005).

1.2 Uppsök, The Undergraduate Theses Service Provider

One of the reasons for choosing to work with undergraduate theses for a national portal is that these publications have not been easily accessible. Work at the universities has often begun with postgraduate theses and research publications, meanwhile undergraduate theses are also much asked for by students and others within a range of subjects. Also, the general public is likely to use this material as an easy way of getting information within a given area. In some professions undergraduate theses are considered a vital and valuable source of information, often in areas where research density is low.

We also argue that, by getting the undergraduate students used to archiving electronic versions of their output at their institutions, already at an early stage of their careers, they are likely to continue to do so later on recognising the benefit of reaching a wider audience with their work and appreciating the benefits of Open Access. In addition,

the institutions may find increased visibility of the results from studies a way of marketing their courses. Another positive factor is that library resources could be used more efficient in this particular area.

During the first year of this project, several universities and university colleges have installed, joined and/or developed systems and solutions for their e-publishing of theses and diploma work. To date, twelve universities have had their metadata harvested for the central service, and the number of data providers is growing. The service, Uppsök, is available since November 2004, at URL <http://www.libris.kb.se/uppsok>.

2 Metadata Modelling

During the early phase of the project, the metadata schemas of local repositories were reviewed in relation to the proposed national schema and a pilot study was tested. Already at that point we found the interest for making this central service was high. The goal was to find a metadata model that primarily could be used for harvesting services, and not for cataloging or in student thesis workflow. The institutions that wanted to participate were required to apply to the metadata model, perform their harvesting according to OAI-PMH protocol and have their theses available in PDF-format. These simple requirements for joining turned out to be very successful.

2.1 Metadata Model for Uppsök

We used the Dublin Core (DC Metadata Element Set, 2005) format as the basis for interoperability in the metadata model for Uppsök. The following elements are mandatory: dc:title, dc:identifier, dc:publisher, dc:type, dc:date, dc:language, dc:creator. Besides the mandatory elements, it is possible to put theses' abstracts in dc:description, and keywords, treated as uncontrolled, and in dc:subject, if applicable. Rules for how to express the semantics in each of these fields are part of the metadata model (Metadata model, 2005), which also includes specific vocabularies, e.g. for dc:type, which describes types of undergraduate theses. To adapt to different kinds of educations in Sweden the theses were subdivided into five levels, eg. types of thesis; A = first term paper, B = second term paper, C = third term thesis, D = fourth term thesis, Y = professional degree thesis. We included papers of lower levels than diploma, bachelor and master degree and at the same time distinguish advanced theses from professional degree theses.

The problem with "degree" in dc:type is that Dublin Core does not have an obvious place for the degree information. The Networked Digital Library of Theses and Dissertations (NDLTD) (ETD-MS, 2005) has in their Metadata Standard added an element "thesis.degree" and defined thesis degree level, "Level of education associated with the document. Examples; bachelor's, master's". UK's FAIR Programme has a Metadata Core Set for ETDs (Electronic Theses, 2005) in which dc:type contains both Qualification level (Diploma, Masters, Doctoral, Postdoctoral ETS and Qualification name (specific degree MPhil, PhD). We should perhaps have followed ETDs metadata model and limited our five "type of thesis" to Qualification level and Qualification name.

We spent much time in the project discussing a common classification of theses types, and from time to time, continue to bring up the subject. It is somewhat a "Gordian knot" and we agree with the recommendation of the DARE project to cut the knot and spare the project endless discussions – "DARE chose the pragmatic solution to work with simple DC and leave more elaborate metadata exchange until such time as it becomes necessary." (Van der Kuil, 2004)

At present, there is no need to revise our dc:type. In the near future it will be necessary to adapt our dc:type definition to a common European higher education standard definition of degree and credit points system following the Bologna Process. The new credit system will be ECTS (European Credit Transfer System). The Swedish credit point system is compatible with the ECTS (one Swedish credit point=1.5 ECTS credit points) but very few Swedish universities have yet decided to use the ECTS points as a general system together with the national credit point system. 40 countries will report current progress in the Bologna Process at a conference in Bergen in May, 2005 (Bologna Process, 2005). The national report from Sweden analyses among many other things, how the proposed changes will affect the Swedish Master's degree (Sweden, 2005).

As part of the metadata model we added a mandatory OAI-Set structure (OAI, 2005) which is divided in eight sets to provide subject-based guidance in the service, but also to act as a base for selective harvesting by other service providers.

Since local repositories use different classification schemes, we decided not to try to implement a common classification or thesaurus. The solution to this problem was to create a set structure that could cover all subjects. The setSpecs are (abbreviated): SocialBehaviourLaw, LifeEarthScience, PhysicsChemistryMaths, Technology, Medicine, AgricultureVeterinaryMedForestry, FineArt. Our experience so far is that these eight groups are satisfactory for searching needs in Uppsök.

3 Advice and Dissemination

The metadata model is the foundation for the national portal. To establish a national service for undergraduate theses with highly informative user value one important goal was to get a majority of Swedish universities and university colleges to join the service. This goal was supported in the SVEP work package for advice and dissemination. We wanted to start up discussions and collaborations and to spread knowledge concerning technical questions and institutional repository software for e-publishing. We also wanted to give help and support to institutions who had just started or were going to start a local repository for e-publishing. The dissemination on e-publishing issues to universities and university colleges was done in following ways:

- By putting together an overview of institutional repository software
- By initiating a workshop series
- By setting up a web site with information and presentations
- By giving help and support by experts to local institutions

3.1 Overview of Institutional Repository Software

Within the project we have put together an overview of institutional repository software (SVEP project homepage – Översikt, 2005), which are compliant with the OAI-PMH and inspired by the excellent Open Society Institute Guide (A Guide to Institutional Repository Software, 2005). The aim of the overview is to help organisations select a software system that best satisfies their institution's needs. The repository software in the overview were either chosen from a selection of free and commercial tools already used in Sweden or from those of special interest in a Swedish context. The overview comprises CDSWare, DiVA, DSpace, EPrints and Lotus Notes and contains a short summary description of each software with contact information and links for further information. The main part of the overview is a feature and functionality table, which provides details on specific system functionality such as status, technical specifications, set-up installations, administration/systems management and metadata/interoperability. With the exception of DiVA and Lotus Notes the systems discussed in the overview are Open Source software allowing some level of local customisation.

3.2 Workshops

Experiences of different publishing tools and other implementation issues were presented at the series of four workshops that took place in different regions of Sweden during one year, from April, 2004- to April, 2005 (SVEP project homepage – Workshops, 2005). The meetings were aimed for dissemination, cooperation and also to recruit new members to the national service for undergraduate theses. Since universities and research centers are actively planning and implementing institutional repositories, the timing for the workshop series was particularly good and the interest was very high. Together the workshops attracted most of the Swedish universities and university colleges and also representatives from other Nordic countries.

Among the subjects covered by the workshops were experiences from local adaptations and questions of how to deal with faculty. We also invited speakers from abroad to offer perspectives and knowledge from work with e-publishing from countries outside Sweden, such as the UK and Norway (SVEP project homepage – Workshops, 2005). The progress with the national portal Uppsök, was presented during the workshop series and was also followed by a competition for finding a “proper” name for the service. Moreover, at one of the workshops the subject of plagiarism was discussed. In a time when we experience a growing amount of students producing a large amount of undergraduate theses, which will be electronically exposed on the Internet, questions about plagiarism are growing.

A web site for the SVEP-project was established to assemble information for users. The overview of e-publishing tools and information about the work within the project together with reports and presentations from the workshops

are presented at the project web site. Our work also provided help support within a network of experts to universities and university colleges in their work on local e-publishing.

4 The Portal Comes Alive

In the early stages of the project we worked with a demonstrator prototype for the search engine and the harvester for which no one had any long term responsibility. But this changed in the summer of 2004 when the LIBRIS Department (responsible for the maintenance and systems development) of the Royal Library in Stockholm, which already participated in the SVEP-project offered to take a long term responsibility for the service (LIBRIS 2005). This gave us on one hand very good conditions but on the other hand limitations in developing the user interface. The positive aspects of this was of course that we had something to work on that was real and could be tested, implemented and used in reality. The negative aspects were that the user interface had to be designed in line with other portals and resources in the LIBRIS system, which to a certain degree limited our possibilities.

In November, 2004 the first version of the interface was released. The portal then contained close to 5000 records from ten universities. As of the twentyseventh of April 2005 we have 6156 records from 12 Universities. Currently the end-user interface is only in Swedish, although circa 25% of the theses are in English.

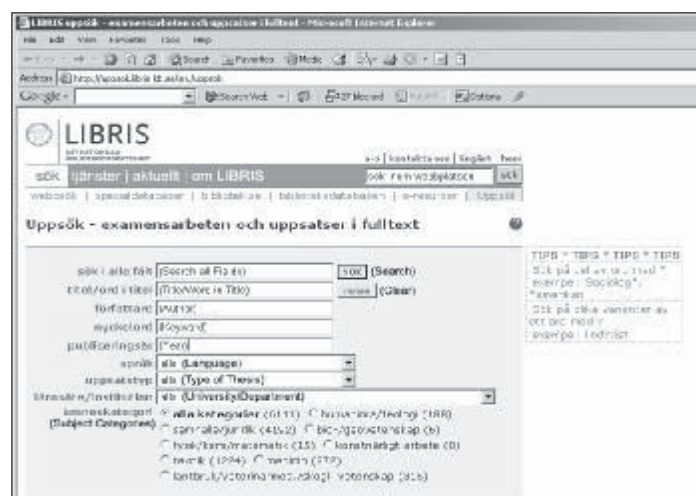


Figure 1. The Uppsök Interface

4.1 The User Study

In order to evaluate the portal, Uppsök, we carried out a user study during January and February, 2005. Using the programme Camtasia Studio (TechSmith Corporation, 2005) we recorded how the user moved about on the screen solving the assigned tasks. The subjects of the user study were seated in front of a computer, they received a short briefing on the content of Uppsök and then were urged to navigate the interface on their own solving six assigned tasks. After every task the subjects were interviewed about their course of action. A form with standard questions such as background, education, computer experience etc. was filled out by every subject of the user study after every session. All the questions asked during the interviews, plus the description of the screen navigation, were transcribed.

10 persons were randomly chosen for the study: 5 women and 5 men. 9 of these were university students, all of them in the final phase of their studies. One of the men was a university lecturer and the only one who was not a student. The six tasks were constructed to give us clues about the following issues:

- How hard/easy is it to navigate the interface?
- Is the field "Type of thesis" useful?
- Is the list of university departments useful or confusing?
- Have we found the right combination of fields and choices for the interface?
- Does the resource contain attractive information?

We chose this qualitative method described above since it allowed us to carry out a comprehensive study that would not comprise extensive interviewing, collection and processing work. The Camtasia Studio programme costs about

300 USD for a single license and it is very easy to use for recording pictures and sounds from a computer. On average, every recording session took 1 hour. This included preparations such as instructions to study subject about the portal and the tasks to be solved. Participants were interviewed and filled out the questionnaire after the recording session. Transcribing every session took about 3 hours. The total time required for the user study totaled about 1 week. Add another week for constructing the set up, creating tasks, questionnaires, writing reports etc.

The result of our study concurs with our earlier experiences of users (students) who present certain homogenous behaviour such as often blindly applying Google searching techniques to any other search service; not using help-facilities; having problems distinguishing between fields such as “Keywords”, “Search in all fields”, and “Subject”. Furthermore the possibility to be able to ask the subjects about their searching behaviour in front of the screen in a quiet surrounding (An office at the library at Blekinge Institute of Technology) and by dialogue try to understand mistakes and accomplished solutions, was very stimulating.

4.2 Results

By creating and realising this kind of user study we think we can safely say that Uppsök offers attractive information in an accessible way for students at universities in Sweden. The portal received very positive remarks from the students – and they appreciated it as a great resource for finding ideas for topics and for learning how to structure their own theses. Nevertheless, the user study points to several shortcomings where measures can be recommended. The most important of these are:

- Divide the field “University” into two choices. First choose “university” and then “department”.
- User instruction must be more highlighted. None of our study subjects found or used the help texts.
- The interface is sometimes confusing. One example is the view with three links to three different search resources placed very near each other.
- The Keyword field is not very helpful. Users have varying conceptions of its meaning and use. Furthermore, there is no certified vocabulary for keywords which makes results less precise.
- The field “Type of Thesis” is of use for the students, even though there is a bit of confusion about the definition, which has to be more clearly explained.

5 Web Statistics

The statistics we have access to are for searches-per-day in Uppsök between November, 2004 and March, 2005. In the Uppsök portal an average of 1004 searches were made every weekday during 14 weeks between November 22nd and March 11th. Two weeks before and after Christmas are subtracted. Saturdays and Sundays are also subtracted since they generally have about 75% lower searching frequency than weekdays. Also on Fridays, one can notice how the search frequency dives to about 30-40% compared with searches done between Monday – Thursday.

The search frequency is evenly distributed between the weeks. So it seems that the marketing efforts, which were pretty limited, still resulted in a steady number of users that returned after the Christmas holiday.

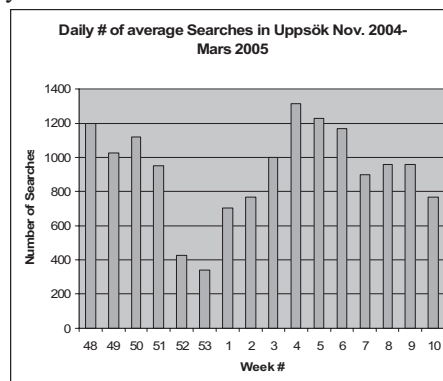


Figure 2. Daily numbers of average searches in Uppsök

6 Conclusions

We believe that with an increasing number of records, and with an increasing number of participating universities and continued marketing, Uppsök will be firmly established as a valuable resource created and used by Swedish

university students. Today 25% of the thesis in Uppsök are in English and we believe that in the future if the system is upgraded with the metadata core set for ETDs it can be even more interesting for students outside Sweden. This project, resulting in a real service with a long term provider and host having over a thousand users every day, tells us that the project was executed in the right place, the right time and with the right content! Several factors also helped to make the project a success: the simplicity of the metadata model (which to our pleasure has influenced the metadata model for open archives in Norwegian university) (Jakobsson, 2004); the project partners were a heterogenous group with great experience in practical work with student thesis and with a common focus; the tools, software and standards were all there for us to use; the workshop series gave a lot of input to the project in the form of new content providers and new ideas and above all – LIBRIS gave the project a stable and recognisable platform for long term commitment.

References

- BIBSAM, the Swedish Royal Library's Department for National Co-ordination and Development: URL <http://www.kb.se/bibsam/english/first.htm>, 050418
- Bologna Process: From Berlin to Bergen, May 2005, <http://www.bologna-bergen2005.no/>, 040517
- DC Metadata Element Set Version 1.1 <http://dublincore.org/documents/2004/12/20/dces/>, 050420
- Electronic Theses, Metadata Core Set for ETDs, The Robert Gordon University Library <http://www2.rgu.ac.uk/library/guidelines/metadata.html>, 050416
- ETD-MS: an Interoperability Metadata Standard for Electronic Theses and Dissertations -- version 1.00, revision 2 <http://www.ndltd.org/standards/metadata/current.html>, 050412
- European Association for International Education, The Bologna process, http://www.eaie.org/about/bologna_eaie.asp, 050413
- A Guide to Institutional Repository Software, Open Society Institute <http://www.soros.org/openaccess/software/>, 050412
- Jakobsson, Arne & Erlandsen, Morten (2004). Forslag til metadatamodell for åpne institusjonelle arkiv ved universitetene i Norge. (in Norwegian only) http://www.ub.uio.no/ubfelles/prosjekter/digipub/Ubena/Metadatamodell/Metadatamodell_for_Norge_version_4.htm, 050425
- LIBRIS web site, <http://www.libris.kb.se>, 050428
- Metadata model for Uppsök: http://www.svep-projekt.se/masters-theses/Metadata_model, 050425
- OAI Protocol for Metadata Harvesting: URL <http://www.openarchives.org/OAI/openarchivesprotocol.html>, 050413
- Open Archives Initiative Protocol for Metadata Harvesting: URL <http://www.openarchives.org/OAI/openarchivesprotocol.html#Set>, 050413
- Sweden: National Report 2004-2005, http://www.bologna-bergen2005.no/EN/national_impl/00_Nat-rep-05/National_Reports-Sweden_050114.pdf, 050420
- SVEP project homepage: URL <http://www.svep-project.se/> 050425
- SVEP project homepage - Översikt URL <http://www.svep-projekt.se/e-publishing/Oversikt/>, 050425
- SVEP project homepage - Workshops URL <http://www.svep-projekt.se/workshops/>, 050525
- TechSmith Corporation. <http://www.techsmith.com/products/studio/default.asp>, 050420
- Van der Kuil, Annemiek & Feijen, Martin (2004). The Dawning of the Dutch Network of Digital Academic Repositories (DARE): A Shared Experience. Ariadne, Issue 41, October, <http://www.ariadne.ac.uk/issue41/vanderkuil/>, 050427

Acknowledgement

We want to thank Jan Hagerlid, BIBSAM. Without his enthusiasm there would not have been any SVEP-project. The LIBRIS staff who gave Uppsök an excellent home and made the right priorities and also provided the statistics. Last but not least, many thanks to Jessica Lindholm, our project manager who led us to basic results before she left us in March 2005 for the greener grass in Denmark.